SUN SPARC® ENTERPRISE T5120 AND T5220 SERVER ARCHITECTURE

Unleashing the UltraSPARC® T2 Processor with CoolThreads™ Technology

White Paper
October 2007

Table of Contents

Executive Summary

Use of the Web is changing in fundamental ways, driven by Web 2.0 applications and the thousands of people who join the global Internet every day through a proliferation of new interactive devices. The character of applications and services is changing too. Increasingly, user's don't need to install anything, upgrade anything, license anything, subscribe to anything, or even buy anything in order to participate and transact. Web users can even interact directly with content, changing it and improving it. Intellectual property is shared, rather than locked away, and the most popular services are available free of charge. Even very small transactions are now encouraged, becoming large in aggregate. Social networking and other collaborative sites let like-minded people from around the world share information on enormous range of topics and issues. Business transactions too are now predominantly Web based.

Serving this dynamic and growing space is becoming very challenging for datacenter operations. Services need to be able to start small and scale very rapidly, often doubling capacity every three months even as they remain highly available. Infrastructure must keep up with these enormous scalability demands, without generating additional administrative burden. Unfortunately, most datacenters are already severely constrained by both real estate and power — and energy costs are rising. There is also a new appreciation for the role that the datacenter plays in reducing energy consumption and pollution. Virtualization has emerged as an extremely important tool as organizations seek to consolidate redundant infrastructure, simplify administration, and leverage under-utilized systems. Security too has never been more important, with increasing price of data loss and corruption. In addressing these challenges, organizations can ill afford proprietary infrastructure that imposes arbitrary limitations.

Employing the UltraSPARC® T2 processor — the industry's first massively threaded system on a Chip (SoC) — Sun SPARC® Enterprise T5120 and T5220 servers offer breakthrough performance and energy efficiency to drive Web 2.0 infrastructure and address other demanding datacenter challenges. Next-generation CoolThreads™ chip multithreading (CMT) technology supports up to 64 threads in as little as one rack unit (RU) — providing increased computational density while staying within variously constrained envelopes of power and cooling. Very high levels of integration help reduce latency, lower costs, and improve security and reliability. Balanced system design provides support for a wide range of application types — from Web services to high performance computing (HPC). Uniformity of management interfaces and adoption of standards helps reduce administrative costs. With both the processor and the Solaris™ Operating System (Solaris OS) available under open source licensing, organizations are free to innovate and join with a world-wide technical community.

Chapter 1

The Evolution of Chip Multithreading (CMT)

By any measure, Sun's first-generation CMT processors were an unprecedented success. Sun Fire™ / Sun SPARC Enterprise T1000 and T2000 servers based on the UltraSPARC T1 processor with CoolThreads technology won enthusiastic praise, and generated the fastest product ramp in Sun's history. Delivering up to five times the throughput in a quarter of the space and power, these systems even garnered the first ever rebate from a power utility[1] — a trend that is being repeated across the world. Now CMT technology is evolving rapidly to meet the constantly changing demands of a wide range of Web and other applications.

Business Challenges for Web 2.0

Marked by the prevalence of Web services and service-oriented architecture (SOA), the emerging *Participation Age* promises the ability to deliver rich new content and high-bandwidth services to larger numbers of users than ever before. Through this transition, organizations across many industries hope to address larger markets, reduce costs, and gain better insights into their customers. At the same time, an increasingly broad array of wired and wireless client devices are bringing network computing into the everyday lives of millions of people. These trends are redefining datacenter scalability and capacity requirements, even as they collide with fundamental real estate, power, and cooling constraints.

- ***Building out for Web Scale***

  Web scale applications engender a new pace and urgency to infrastructure deployment. Organizations must accelerate time to market and time to service, while delivering scalable high-quality and high-performance applications and services. Many need to be able to start small with the ability to scale very quickly, with new customers and innovative new Web services often implying a doubling of capacity in months rather than years.

  At the same time, organizations must reduce their environmental impact by working within the power, cooling, and space available in their current datacenters. Operational costs too are receiving new scrutiny, along with system administrative costs that can account for up to 40 percent of an IT budget. Simplicity and speed are paramount, giving organizations the ability to respond quickly to dynamic business conditions. Organizations are also striving to eliminate vendor lock-in as they look to preserve previous, current, and future investments. Open platforms built around open standards help provide maximum flexibility while reducing costs of both entry and exit.

1.In August of 2006, Pacific Gas and Electric (PG&E) began offering a substantial energy rebate for purchasing and deploying Sun Fire / Sun SPARC Enterprise T1000 and T2000 servers

- ***Driving Datacenter Virtualization and Eco-Efficiency***
  Coincident with the need to scale services, many datacenters are recognizing the advantages of deploying fewer standard platforms to run a mixture of commercial and technical workloads. This process involves consolidating under-utilized and often sprawling server infrastructures with effective virtualization solutions that serve to enhance business agility, improve disaster recovery, and reduce operating costs. This focus can help reduce energy costs and break through datacenter capacity constraints by improving the amount of realized performance for each watt of power the datacenter consumes.

  Eco-efficiency provides tangible benefits, improving *ecology* by reducing carbon footprint to meet legislative and corporate social responsibility goals, even as it improves the *economy* of the organization paying the electric bill. As systems are consolidated onto more dense and capable computing infrastructure, demand for datacenter real estate is also reduced. With careful planning, this approach can also improve service uptime and reliability by reducing hardware failures resulting from excess heat load. Servers with high levels of standard reliability, availability, and serviceability (RAS) are now considered a requirement.

- ***Securing the Enterprise at Speed:***
  Organizations are increasingly interested in securing all communications with their customers and partners. Given the risks, end-to-end encryption is essential in order to inspire confidence in security and confidentiality. Encryption is also increasingly important for storage, helping to secure stored and archived data even as it provides a mechanism to detect tampering and data corruption.

  Unfortunately, the computational costs of increased encryption can increase the burden on already over-taxed computational resources. Security also needs to take place at line speed, without introducing bottlenecks that can impact the customer experience or slow transactions. Solutions must help to ensure security and privacy for clients and bring business compliance for the organization, all without impacting performance or increasing costs.

Rule-Changing Chip Multithreading (CMT) Technology

Addressing these challenges has outstripped the capabilities of traditional processors and systems, and required a fundamentally new approach.

Moore's Law, and the Diminishing Returns of Traditional Processor Design

The oft-quoted tenant of Moore's Law states that the number of transistors that will fit in a square inch of integrated circuitry will approximately double every two years. For over three decades the pace of Moore's law has held, driving processor performance to new heights. Processor manufacturers have long exploited these gains in chip real

estate to build increasingly complex processors, with instruction-level parallelism (ILP) as a goal. Today these traditional processors employ very high frequencies along with a variety of sophisticated tactics to accelerate a single instruction pipeline, including:

- Large caches
- Superscalar designs
- Out-of-order execution
- Very high clock rates
- Deep pipelines
- Speculative pre-fetches

While these techniques have produced faster processors with impressive-sounding multiple-gigahertz frequencies, they have largely resulted in complex, hot, and power-hungry processors that are not well suited to the types of workloads often found in modern datacenters. In fact, many datacenter workloads are simply unable to take advantage of the hard-won ILP provided by these processors. Applications with high shared memory and high simultaneous user or transaction counts are typically more focused on processing a large number of simultaneous threads (thread-level parallelism, TLP) rather than running a single thread as quickly as possible (ILP).

Making matters worse, the majority of ILP in existing applications has already been extracted and further gains promise to be small. In addition, microprocessor frequency scaling itself has leveled off because of microprocessor power issues. With higher clock speeds, each successive processor generation has seemingly demanded more power than the last, and microprocessor frequency scaling has leveled off in the 2-3 GHz range as a result. Deploying pipelined Superscalar processors requires more power, limiting this approach by the fundamental ability to cool the processors.

Chip Multiprocessing with Multicore Processors

To address these issues, many in the microprocessor industry have used the transistor budget provided by Moore's Law to group two or even four conventional processor cores on a single physical die — creating multicore processors (or chip multiprocessors, CMP). The individual processor cores introduced by many CMP designs have no greater performance than previous single-processor chips, and in fact, have been observed to run single-threaded applications more slowly than single-core processor versions. However, the aggregate chip performance increases since multiple programs (or multiple threads) can be accommodated in parallel (thread level parallelism).

Unfortunately, most currently-available (or soon to be available) chip multiprocessors simply replicate cores from existing (single-threaded) processor designs. This approach typically yields only slight improvements in aggregate performance since it ignores key performance issues such as memory speed and hardware thread context switching. As

a result, while these designs provide some additional throughput and scalability, they can consume considerable power and generate significant heat — without a commensurate increase in overall performance.

Chip Multithreading (CMT) with CoolThreads™ Technology

Sun engineers were early to recognize the disparity between processor speeds and memory access rates. While processor speeds continue to double every two years, memory speeds have typically doubled only every six years. As a result, memory latency now dominates much application performance, erasing even very impressive gains in clock rates. This growing disconnect is the result of memory suppliers focusing on density and cost as their design center, rather than speed.

Unfortunately, this relative gap between processor and memory speeds leaves ultra-fast processors idle as much as 85 percent of the time, waiting for memory transactions to complete. Ironically, as traditional processor execution pipelines get faster and more complex, the effect of memory latency grows — fast, expensive processors spend more cycles doing nothing. Worse still, idle processors continue to draw power and generate heat. It is easy to see that frequency (gigahertz) is truly a misleading indicator of real performance.

First introduced with the UltraSPARC T1 processor, chip multithreading takes advantage of CMP advances, but adds a critical capability — *the ability to scale with threads rather than frequency*. Unlike traditional single-threaded processors and even most current multicore (CMP) processors, hardware multithreaded processor cores allow rapid switching between active threads as other threads stall for memory. Figure 1 illustrates the difference between CMP, fine-grained hardware multithreading (FG-MT), and chip multithreading. The key to this approach is that each core in a CMT processor is designed to switch between multiple threads on each clock cycle. As a result, the processor's execution pipeline remains active doing real useful work, even as memory operations for stalled threads continue in parallel.

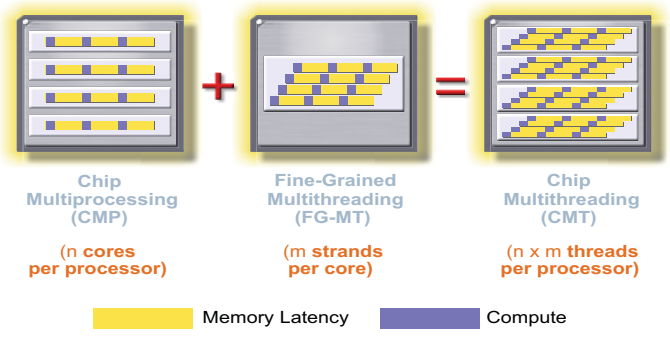

Figure 1. Chip multithreading combines CMP and fine-grained hardware multithreading

Chip multithreading provides real value since it increases the ability of the execution pipeline to do actual work on any given clock cycle. Utilization of the processor pipeline is greatly enhanced since a number of execution threads now share its resources. The negative effects of memory latency are effectively masked, since the processor and memory subsystems remain active in parallel to the processor execution pipeline. Because these individual processor cores implement much simpler pipelines that focus on scaling with threads rather than frequency (emphasizing TLP over ILP), they are also substantially cooler and require significantly less electrical energy to operate. This innovative approach results in CoolThreads processor technology — multiple physical instruction execution pipelines (one for each core), with multiple active thread contexts per core.

The UltraSPARC® T2 Processor with CoolThreads Technology

Unlike complex single-threaded processors, CMT processors utilize the available transistor budget to implement multiple hardware multithreaded processor cores on a chip die. The UltraSPARC T2 processor takes the CMT model to the next level, providing up to eight cores with each core supporting up to eight threads via two independent pipelines per core — effectively doubling the throughput of the UltraSPARC T1 processor. In addition, the UltraSPARC T2 processor uses this transistor budget to implement the industry's first massively threaded System on a Chip (SoC), with a single processor die hosting:

- Up to 64 threads per processor (up to eight cores supporting eight threads each)
- On-chip Level-1 and Level-2 caches
- Per-core floating point capabilities
- Per-core cryptographic acceleration
- Two on-chip 10 Gb Ethernet interfaces
- On-chip PCI Express interface

Through this SoC design, the UltraSPARC T2 processor significantly enhances the general purpose nature of the CPU by building in eight floating point units (1 per core). Enhanced floating point capabilities open the UltraSPARC T2 to the world of compute-intensive applications as well as the traditionally CMT friendly datacenter throughput applications. No-cost security and cryptographic acceleration is provided by the on-chip, per-core streaming accelerators. In addition, the ability to move data in and out of the processor is significantly aided by an integrated PCI-Express interface and dual 10 Gigabit Ethernet interfaces.

Sun SPARC® Enterprise T5120 and T5220 Servers

Sun SPARC Enterprise T5120 and T5220 servers (Figure 2) are designed to leverage the considerable resources of the UltraSPARC T2 processor in the form of cost-effective general-purpose platforms. These systems deliver up to twice the throughput of their

predecessors, while leading competitors in terms of performance, performance per watt, and *SWaP* performance (as evaluated by the Space, Watts, and Performance metric detailed later in this section). These systems also extend the benefits of CMT from multithreaded commercial workloads into technical workloads rich in floating point operations.

With support for 64 threads, large memory, cryptographic acceleration, and integrated on-chip 10 Gb Ethernet networking and I/O technology, these servers represent a departure from traditional system design. The 1U Sun SPARC Enterprise T5120 is ideal for providing high throughput within significant power, cooling, and space constraints, delivering a 64-thread UltraSPARC T2 processor in a space- and power-efficient 1U rackmount package. As a compute node with massively horizontally scaled environments, the Sun SPARC Enterprise T5120 can help provide a substantial building block for application tier, Web services, or even high performance computing (HPC) infrastructure. Network infrastructure applications such as portal, directory, network identity, file service, and backup are all a good fit for this server.

The Sun SPARC Enterprise T5220 server provides both throughput as well as expandability, with extra I/O and internal disk options afforded by the 2U rackmount form factor. With greater I/O and internal disk, typical workloads include demanding mid-tier application server deployments or Web- and application-tier consolidation, virtualization, and consolidation projects requiring maximum uptime with future growth and integration into diverse environments. The Sun SPARC Enterprise T5220 is also ideal for OLTP database deployments.

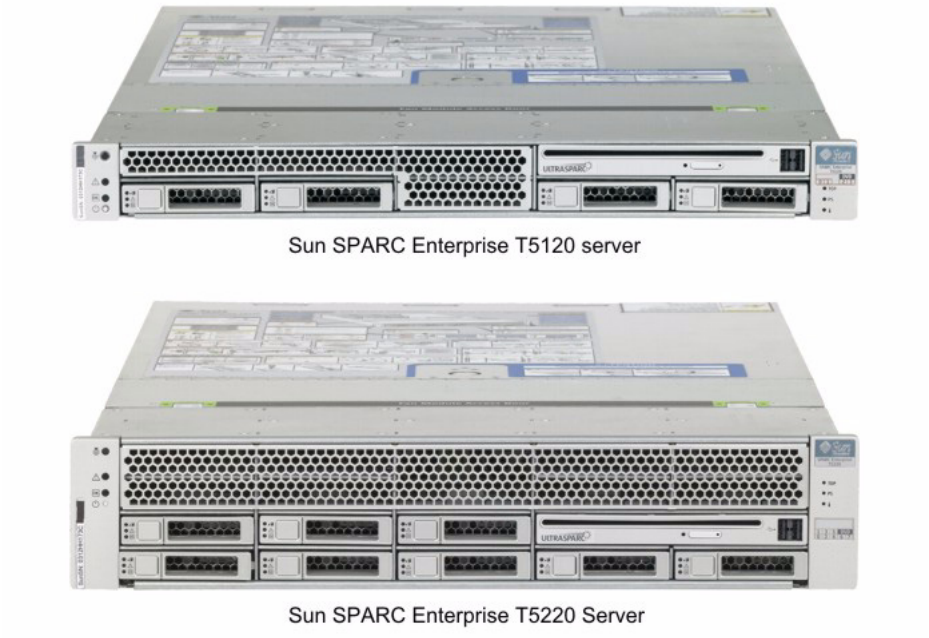

Figure 2. Sun SPARC Enterprise T5120 and T5220 servers

Designed to complement each other, as well as the rest of Sun's server product line, the Sun SPARC Enterprise T5120 and T5220 servers address the dynamic needs of the modern datacenter.

- ***Efficient and Predictable Scalability***
  With support for 64 threads and large memories, Sun SPARC Enterprise T5120 and T5220 servers are the first to utilize the 10 Gb Ethernet, I/O, and cryptographic acceleration provided directly by the UltraSPARC T2 processor chip. This approach provides leading levels of performance and scalability with extremely high levels of power, heat, and space efficiency.

- ***Accelerated Time to Market***
  Sun SPARC Enterprise T5120 and T5220 servers running the Solaris OS provide full binary compatibility with earlier UltraSPARC systems, preserving investments and speeding time to market. Sun's CoolTools for SPARC help accelerate application selection, profiling, testing, tuning, debugging and deployment of key applications on CMT systems.

- ***Simplified Management***
  Each Sun SPARC Enterprise T5120 and T5220 server provides an Integrated Lights Out Management (ILOM) service processor, compatible with Sun's x64 servers. ILOM provides a command line interface (CLI), a Web-based graphical user interface (GUI), and Intelligent Platform Management Interface (IPMI) functionality to aid with out-of-band monitoring and administration. ILOM on these systems also provides an Advanced Lights Out Management (ALOM) backwards compatibility mode for administrators familiar with Sun Fire / Sun SPARC Enterprise T1000 and T2000 servers.

- ***The Industry's Most Open Platform***
  Sun SPARC Enterprise servers are the industry's most open platforms, providing the only mainstream processor and hypervisor offered under the GNU General Public License (GPL). These systems offer a choice of operating systems, including the Solaris OS, Linux, and BSD variants. The Solaris OS is free and open, offering full binary compatibility and enterprise-class features. Sun Java™ Enterprise System middleware is pre-loaded.

- ***Industry-Leading Tools for Virtualization and Consolidation***
  Sun's chip multithreading (CMT) technology is ideal for consolidation, providing low-level multithreading support for virtualization at every layer of the technology stack. Sun Logical Domains exploit the UltraSPARC T2 processor's 64 threads to support multiple guest operating system instances, while Solaris Containers provide virtualization within a single Solaris OS instance. The advanced ZFS file system provides storage virtualization for storage and considerable scalability.

- ***A Tradition of Leading Eco Efficiency***
  Sun Fire / Sun SPARC Enterprise T1000 and T2000 servers were the industry's first eco-responsible servers. Sun SPARC Enterprise T5120 and T5220 servers continue this tradition by offering the best performance and performance-per-watt across a wide range of commercial and technical workloads. In addition, the UltraSPARC T2 processor is the first and only processor to incorporate unique power management features at both core and memory levels of the processor.
- ***System and Datacenter Reliability***
  Reliability is key to keeping applications available and costs down. With the greater levels of integration provided by an SoC design, the Sun SPARC Enterprise T5120 and T5220 servers offer greatly reduced part counts, and provide commensurately higher levels of reliability, availability, and serviceability (RAS). Lower power consumption and higher performance per watt greatly reduce generated heat loads and the associated issues they cause. Technologies such as Solaris Predictive Self Healing are integrated with the hardware, and help keep systems available.
- ***Zero-Cost Security***
  Providing secure communications and data protection has never been more important, with attempted electronic intrusion and theft at an all-time high. With up to eight integrated cryptographic accelerators on each UltraSPARC T2 processor, there is simply no need to send plain text on the network or store plain text in storage systems. Sun SPARC Enterprise T5120 and T5220 servers support many more crypto operations per second than competitive processors and dedicated crypto accelerator cards, with minimal impact to system overhead.

Table 1 compares the features of Sun SPARC Enterprise T5120 and T5220 servers.

Table 1. Sun SPARC Enterprise T5120 and T5220 server features

Feature	Sun SPARC Enterprise T5120 Server	Sun SPARC Enterprise T5220 Server
CPUs	Four-, six-, or eight-core 1.2 GHz or 1.4 GHz UltraSPARC T2 processor	Four-, six-, or eight-core 1.2 GHz or eight-core 1.4 GHz UltraSPARC T2 processor
Threads	Up to 64	Up to 64
Memory capacity	Up to 64 GB (1, 2, or 4 GB FBDIMM)	Up to 64 GB (1, 2, of 4 GB FBDIMM)
Maximum internal disk drives	Up to four SFF 2.5-inch SAS 73 or 146 GB disk drives, RAID 0/1	Up to eight SFF SAS 2.5-inch SAS 73 or 146 GB disk drives, RAID 0/1
Removable/pluggable I/O	Slimline DVD-R Four USB 2.0 ports	Slimline DVD-R Four USB 2.0 ports
PCI	One x8 PCI Express slot, Two x4 PCI Express or XAUI combo slots[a]	Two x8 PCI Express Two x4 PCI Express Two x4 PCI Express or XAUI combo slots[a]

Feature	Sun SPARC Enterprise T5120 Server	Sun SPARC Enterprise T5220 Server
Ethernet	Four on-board Gigabit Ethernet ports (10/100/1000) Two 10 Gb Ethernet ports via XAUI combo slots	Four on-board Gigabit Ethernet ports (10/100/1000) Two 10 Gb Ethernet ports via XAUI combo slots
Power supplies	Two hot-swappable AC 650W or DC 660W power supply units (N+1 redundancy)	Two hot-swappable AC 750W power supply units (N+1 redundancy)
Fans	Four hot-swappable fan trays, with 2 fans per tray, N+1 redundancy	3 hot-swappable fan trays, with 2 fans per tray, N+1 redundancy
Form factor	1 rack unit (1U)	2 rack units (2U)

a.Optional XAUI adapter cards required for access to dual 10 Gb Ethernet ports on the UltraSPARC T2 processor

Space, Watts, and Performance: Introducing the SWaP Metric

Sun SPARC Enterprise T5120 and T5220 servers deliver leading performance across a range of multithreaded workloads and benchmarks. However, with energy and real estate costs and pressures, it is not enough to measure performance in isolation. Delivering the required level of throughput in a fixed space and power envelope is critical. Traditional system-to-system benchmarks are valuable as a way of comparing one system to another, but are limited when it comes to understanding the power and density attributes of the systems being compared. For this reason, Sun has developed the SWaP metric, standing for *Space, Watts, and Performance*. Designed to provide a simple and transparent measure of overall server efficiency, SWaP is calculated using the following formula:

SWaP = Performance / (Space * Power Consumption) where,

- *Performance* is measured by industry-standard audited benchmarks such as those sponsored by the Systems Performance Evaluation Corporation (SPEC)
- *Space* refers to the height of the server in rack units (RUs)
- *Power* is measured by watts used by the system, taken during actual benchmark runs or from vendor's site planning guides

For the latest SWaP results for a variety of benchmarks, please see *sun.com/servers/coolthreads/benchmarks*.

Chapter 2

The UltraSPARC T2 Processor with CoolThreads Technology

The UltraSPARC T2 processor is the industry's first system on a chip (SoC), supplying the most cores and threads of any general-purpose processor available, and integrating all key system functions.

UltraSPARC T2: The World's First Massively Threaded System on a Chip (SoC)

The UltraSPARC T2 processor eliminates the need for expensive custom hardware and software development by integrating computing, networking, security, and I/O on to a single chip. Binary compatible with earlier UltraSPARC processors, no other processor delivers so much performance in so little space and with such small power requirements — letting organizations rapidly scale the delivery of new network services with maximum efficiency and predictability. The UltraSPARC T2 processor is shown in Figure 3, to the left of the previous-generation UltraSPARC T1 processor. Even with twice the computational throughput and significantly higher levels of integration, the UltraSPARC T2 processor is physically smaller than the UltraSPARC T1 processor.

Figure 3. The UltraSPARC T2 (left) and UltraSPARC T1 Processors with CoolThreads Technology

Each UltraSPARC T2 processor supports:

- Up to eight cores @ 1.2 Ghz – 1.4 Ghz
- Eight threads per core for a total maximum of 64 threads per processor
- 4 MB L2 cache in eight banks (16-way set associative)
- Four on-chip memory controllers for support of up to 16 FBDIMMs
- Up to 64 GB of memory (4 GB FBDIMMs) with 60 GB/s memory bandwidth
- Eight fully pipelined floating point units (1 per core)
- Dual on-chip 10 Gb Ethernet interfaces
- Integral PCI-Express interface
- Eight stream processing units (cryptographic co-processors), one per core

Table 2 provides a comparison between the UltraSPARC T2 and the UltraSPARC T1 processor.

Table 2. UltraSPARC T1 and T2 processor features

Feature	UltraSPARC T1 Processor	UltraSPARC T2 processor
Cores per processor	Up to 8	Up to 8
Threads per core	4	8
Threads per processor	32	64
Hypervisor	Yes	Yes
Memory	4 memory controllers, 4 DIMMs per controller	4 memory controllers, up to 64 FBDIMMs
Caches	16 KB instruction cache, 8 KB data cache, 3 MB L2 cache (4 banks, 12-way associative)	16 KB instruction cache, 8 KB data cache, 4 MB L2 cache (8 banks, 16-way associative)
Technology	9-layer Cu metal, CMOS process, 90 nm technology	65 nm technology
Floating point	1 FPU per chip	1 FPU per core, 8 FPUs per chip
Integer resources	Single execution unit	2 integer execution units per core
Cryptography	Accelerated modular arithmetic operations (RSA)	Stream processing unit per core, support for the 10 most popular ciphers
Additional on-chip resources	—	Dual 10 Gb Ethernet interfaces, PCI Express interface (x8)

Taking Chip Multithreaded Design to the Next Level

When Sun's in-house design team set out to design the next-generation of CMT processors, they started with key goals in mind:

- Increasing computational capabilities to meet the growing demand from Web applications by providing twice the throughput of the UltraSPARC T1 processor
- Supporting larger and more diverse workloads with greater floating point performance
- Powering faster networking to serve new network-intensive content
- Providing end-to-end datacenter encryption
- Increasing service levels and reducing downtime
- Improving datacenter capacities while reducing costs

CMT architecture is ultimately very flexible, allowing different modular combinations of processors, cores, and integrated components. The considerations listed above drove an internal engineering effort that compared different approaches with regard to making improvements on the successful UltraSPARC T1 architecture. For example,

simply increasing the number of cores would have gained additional throughput, but would have resulted in consuming extra die area, leaving no room for integrated components such as floating point processors.

The final UltraSPARC T2 processor design recognizes that memory latency is truly *the* bottleneck to improving performance. By increasing the number of threads supported by each core, and by further increasing network bandwidth, the UltraSPARC T2 is able provide approximately twice the throughput of the UltraSPARC T1 processor. Each UltraSPARC T2 processor provides up to eight cores, with each core able to switch between up to eight threads (64 threads per processor). In addition, each core provides two integer execution units, so that a single UltraSPARC core is capable of executing two threads at a time. Figure 4 provides a simplified high-level illustration of the thread model supported by an eight-core UltraSPARC T2 processor.

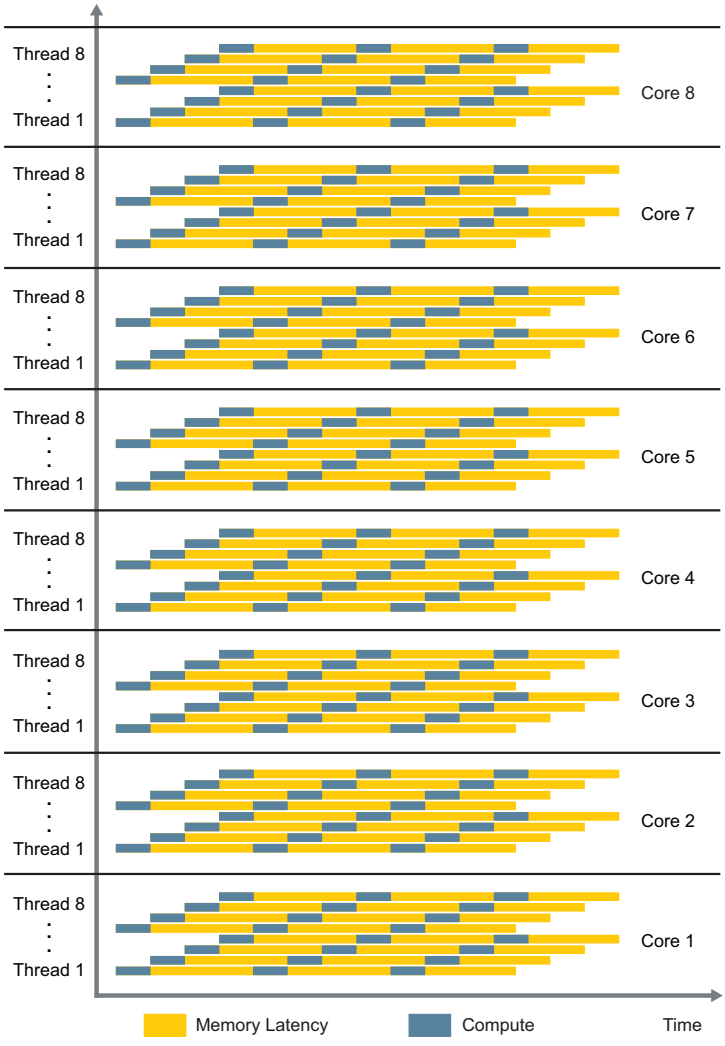

Figure 4. A single eight-core UltraSPARC T2 processor supports up to 64 threads, with up to two threads running in each core simultaneously

UltraSPARC T2 Processor Architecture

The UltraSPARC T2 processor extends Sun's Throughput Computing initiative with an elegant and robust architecture that delivers real performance to applications. A high-level block diagram of the UltraSPARC T2 processor is shown in Figure 5.

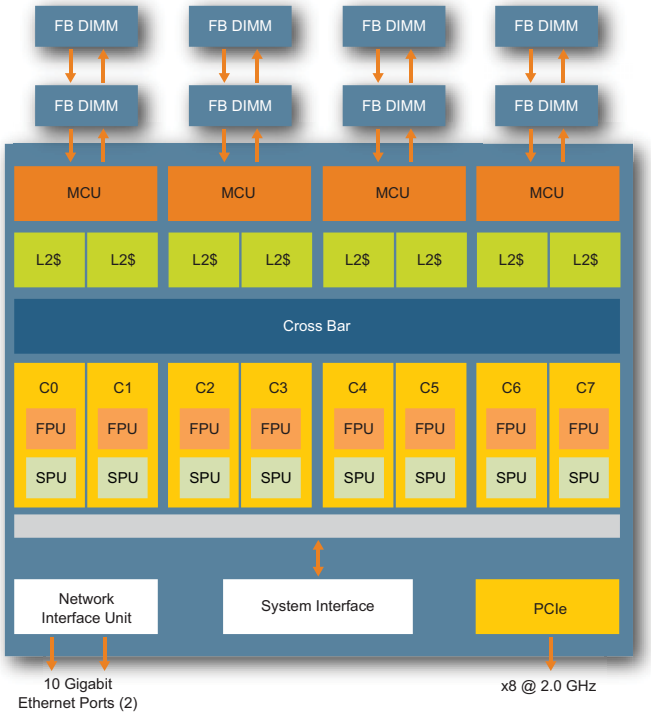

Figure 5. The UltraSPARC T2 processor combines eight cores, memory management, cryptographic support, 10 Gb Ethernet, and PCI Express on a single chip

The eight cores on the UltraSPARC T2 processor are interconnected with a full on-chip non-blocking 8 x 9 crossbar switch. The crossbar connects each core to the eight banks of L2 cache, and to the system interface unit for IO. The crossbar provides approximately 300 GB/second of bandwidth and supports 8-byte writes from a core to a bank and 16-byte reads from a bank to a core. The system interface unit connects networking and I/O directly to memory through the individual cache banks. Using FBDIMM memory supports dedicated northbound and southbound lanes to and from the caches to accelerate performance and reduce latency. This approach provides higher bandwidth than with DDR2 memory, with up to 42.4 GB/second of read bandwidth and 21 GB/second of write bandwidth.

Each core provides its own fully-pipelined floating point and graphics unit (FGU), as well as a stream processing unit (SPU). The FGUs greatly enhance floating point performance over that of the UltraSPARC T1, while the SPUs provide wire-speed cryptographic acceleration with over 10 popular ciphers supported, including DES,

3DES, AES, RC4, SHA-1, SHA-256, MD5, RSA to 2048 key, ECC, and CRC32. Embedding hardware cryptographic acceleration for these ciphers allows end-to-end encryption with no penalty in either performance or cost.

UltraSPARC T2 Core Architecture and Pipelines

Figure 6 provides a block-level diagram representing a single UltraSPARC cores on the UltraSPARC T2 processor (up to eight are supported per processor).

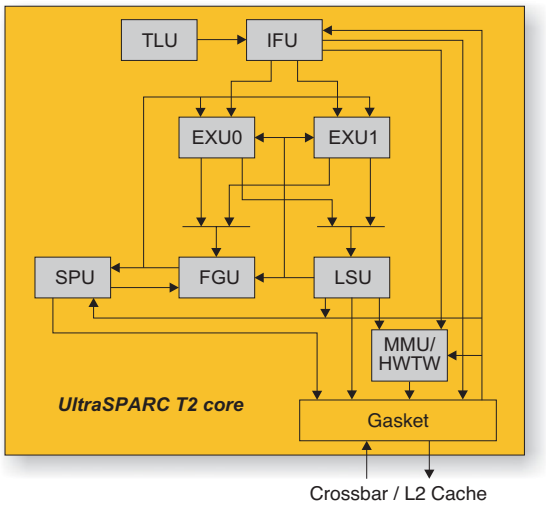

Figure 6. UltraSPARC T2 core block diagram

Components implemented in each core include:

- *Trap Logic Unit (TLU)* — The trap logic unit updates the machine state as well as handling exceptions and interrupts.
- *Instruction Fetch Unit (IFU)* — The instruction fetch unit includes the 16KB instruction cache (32-byte lines, 8-way set associative) and a 64-entry fully-associative instruction translation lookup buffer (ITLB).
- *Integer Execution Units (EXU)* — Dual integer execution units are provided per core with four threads sharing each unit. Eight register windows are provided per thread, with 160 integer register file (IRF) entries per thread.
- *Floating Point/Graphics Unit (FGU)* — A floating point/graphics unit is provided within each core and it is shared by all eight threads assigned to the core. 32 floating-point register file entries are provided per thread.
- *Stream Processing Unit (SPU)* — Each core contains a stream processing unit that provides cryptographic coprocessing.
- *Memory Management Unit (MMU)* — The memory management unit provides a hardware table walk (HWTW) and supports 8 KB, 64 KB, 4 MB, and 256 MB pages.

An eight-stage integer pipeline and a 12-stage floating-point pipeline are provided by each UltraSPARC processor core (Figure 7). A new "pick" pipeline stage has been added to choose two threads (out of the eight possible per core) to execute each cycle.

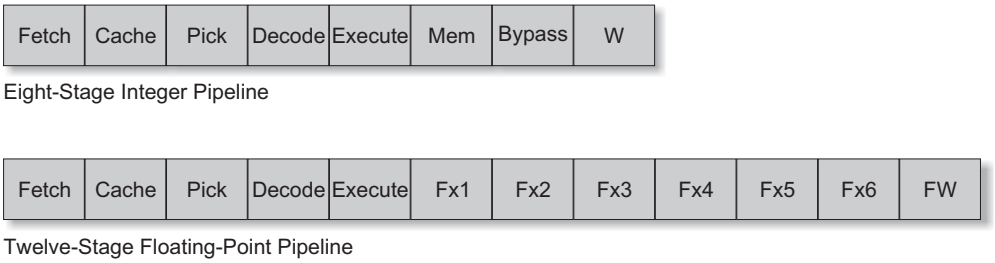

Figure 7. UltraSPARC T2 per-core integer and floating-point pipelines

To illustrate how the dual pipelines function, Figure 8 depicts the integer pipeline with the load store unit (LSU). The instruction cache is shared by all eight threads within the core. A least-recently-fetched algorithm is used to select the next thread to fetch. Each thread is written into a thread-specific instruction buffer (IB) and each of the eight threads is statically assigned to one of two thread groups within the core.

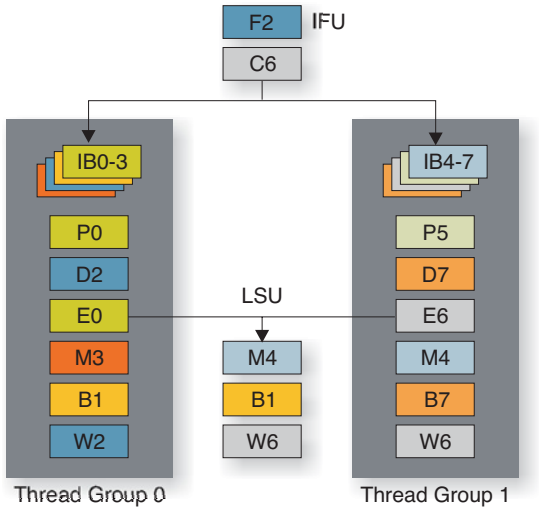

Figure 8. Threads are interleaved between pipeline stages with very few restrictions (integer pipeline shown, letters depict pipeline stages, numbers depict different scheduled threads)

The "pick" stage chooses one thread each cycle within each thread group. Picking within each thread group is independent of the other, and a least-recently-picked algorithm is used to select the next thread to execute. The decode state resolves resource conflicts that are not handled during the pick stage. As shown in the illustration, threads are interleaved between pipeline stages with very few restrictions. Any thread can be at the fetch or cache stage, before being split into either of the two thread groups. Load/store and floating point units are shared between all eight threads. Only one thread from either thread group can be scheduled on such a shared unit.

Integrated Networking

By providing integrated on-chip networking, the UltraSPARC T2 processor is able to provide better networking performance. All network data is supplied directly from and to main memory. Placing networking so close to memory reduces latency, provides higher memory bandwidth, and eliminates inherent inefficiencies of I/O protocol translation.

The UltraSPARC T2 processor provides two 10 Gb Ethernet ports with integrated serdes, offering line-rate packet classification at up to 30 million packets/second (based on layers 1-4 of the protocol stack). Multiple DMA engines (16 transmit and 16 receive DMA channels) match DMAs to individual threads, providing binding flexibility between ports and threads. Virtualization support includes provisions for eight partitions, and interrupts may be bound to different hardware threads.

Stream Processing Unit

The stream processing unit on each UltraSPARC T2 core runs in parallel with the core at the same frequency. Two independent sub-units are provided along with a DMA engine that shares the core's crossbar port:

- A Modular Arithmetic Unit (MAU) shares the FGU multiplier, providing RSA encryption/decryption, binary and integer polynomial functions, as well as elliptic curve cryptography (ECC)[1]
- The cipher/hash unit provides support for popular RC4, DES/3DES, AES-128/192/256, MD5, SHA-1, and SHA-256 ciphers

The SPU is designed to achieve wire-speed encryption and decryption on both of the processor's 10 GB Ethernet ports.

Integral PCI Express Support

The UltraSPARC T2 processor provides an on-chip PCI Express interface that operates at 4 GB/second bidirectionally through a point-to-point dual-simplex chip interconnect. An integral IOMMU supports I/O virtualization and process device isolation by using the PCI Express BDF number. The total I/O bandwidth is 3-4 GB/second, with maximum payload sizes of 128 to 512 bytes. An x8 serdes interface is provided for integration with off-chip PCI Express switches.

1.Supported in a future Solaris OS release

Power Management

Beyond the inherent efficiencies of CMT design, the UltraSPARC T2 is the first processor to incorporate unique power management features at both the core and memory levels of the processor. These features include reduced instruction rates, parking of idle threads and cores, and ability to turn off clocks in both cores and memory to reduce power consumption. Substantial innovation is present in the areas of:

- Limiting speculation such as conditional branches not taken
- Extensive clock gating in the data path, control blocks, and arrays
- Power throttling that allows extra stall cycles to be injected into the decode stage

Chapter 3

Sun SPARC Enterprise T5120 and T5220 Server Architecture

Sun SPARC Enterprise T5120 and T5220 servers have been designed to provide breakthrough performance while maximizing reliability and minimizing power consumption and complexity. This section details the physical and architectural aspects of these systems.

System-Level Architecture

A unified motherboard design is common to both the Sun SPARC Enterprise T5120 and T5220 servers (Figure 9). The motherboard is a 20-layer printed circuit board (PCB) containing the UltraSPARC T2 processor, FBDIMM sockets for main memory, ILOM service processor, disk controller, and I/O subsystems. I/O options include USB, DVD control, quad Gigabit Ethernet, and two levels of PLX PCI Express branching out into sockets for a wide variety of third-party PCI Express expansion options. Shaded regions indicate features that are only available on the Sun SPARC Enterprise T5220 server.

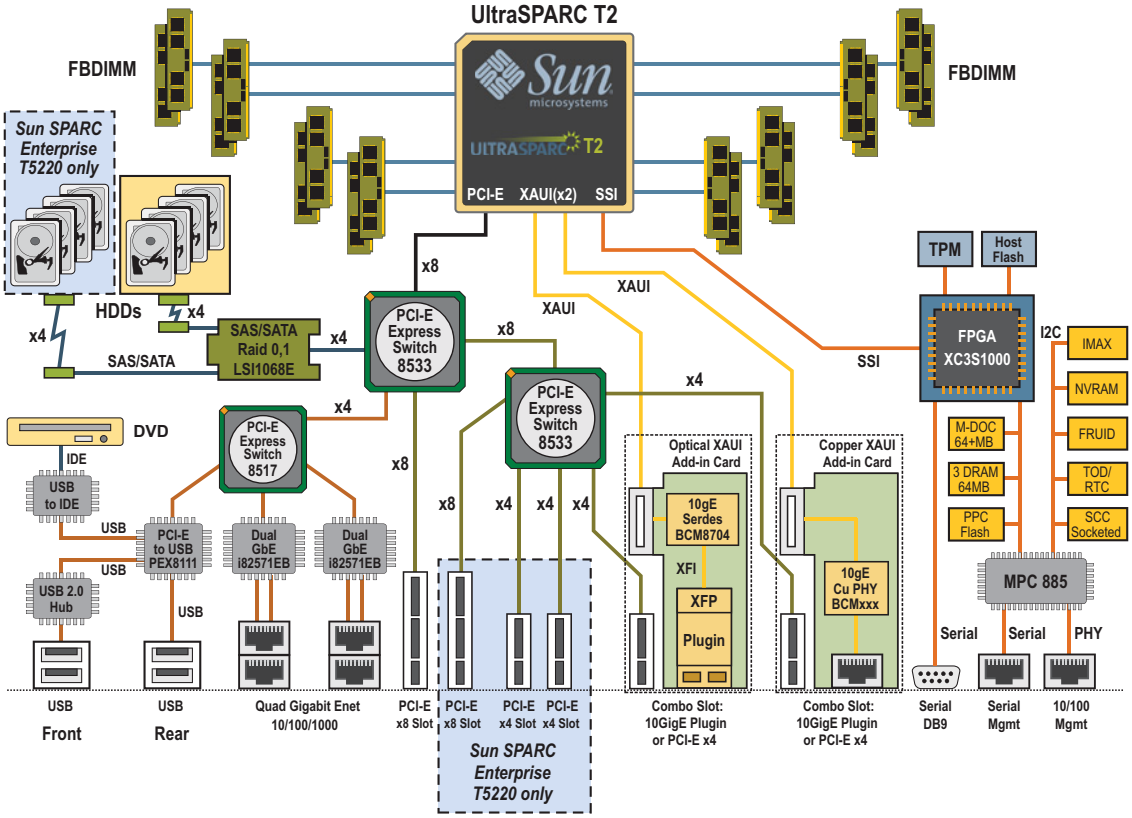

Figure 9. Block-level diagram of the common Sun SPARC Enterprise T5120/T5220 server motherboard

The motherboard interconnect for these systems has been greatly simplified. 12-volt power is distributed to the motherboard through a pair of metal bus bars, connected to a Power Distribution Board (PDB). A single flex-circuit connector routes all critical power control and DVD drive signaling over to the PDB. One or two mini-SAS cables connect the motherboard to the disk drive backplane, providing data access to the system hard drives.

Memory Subsystem

In Sun SPARC Enterprise T5120 and T5220 servers, the UltraSPARC T2 processor's on-chip memory controllers communicate directly to FBDIMM memory through high-speed serial links. The four dual-channel FBDIMM memory controllers can transfer data at an aggregate rate of 32 giga-transfers per second. Sixteen memory socket locations provide sufficient board space for two rows of 667 MHz FBDIMMs per channel.

I/O Subsystem

The UltraSPARC T2 processor incorporates a single, 8-lane (x8) PCI Express port capable of operating at 4 GB/second bidirectionally. This port natively interfaces to the I/O devices through a series of PLX technology PCI Express expander chips, connecting either to PCI Express card slots, or to bridge devices that interface with PCI Express, such as those listed below.

- *Disk Controller* — Disk control is managed by a single LSI Logic SAS1068E SAS controller chip that interfaces to a four-lane (x4) PCI Express port. RAID levels 0 and 1 are provided as standard.
- *Dual GBE* — Two x4 PCI Express ports connect to two Intel Ophir dual Gb Ethernet chips, providing four 10/100/1000 Mbps Ethernet interfaces on the rear of each chassis.
- *USB and DVD* — A single-lane PCI Express port connects to a PLX PEX8111 PCI bridge device. A second bridge chip converts the 32-bit 33MHz PCI bus into multiple USB 2.0 ports. The system's USB interconnect is driven from those ports. In addition, the DVD is driven from a further bridge chip that interfaces one of the USB ports to IDE format.

To minimize cabling and increase reliability, a variety of smaller boards and riser cards are deployed, appropriate to each chassis. These infrastructure boards serve various functions in the Sun SPARC Enterprise T5120 and T5220 servers.

- Power distribution boards distribute system power from the dual power supplies to the motherboard and to the disk backplane (via a connector board)
- Connector boards eliminate the need for many discrete cables, providing a direct card plug-in interconnect to distribute control and most data signals to the disk backplane, fan boards, and the PDB.

- Fan boards provide connections for power and control for both the primary and secondary fans in the front of the chassis. No cables are required since every dual fan module plugs directly into one of these PCBs which, in turn, plugs into the Connector Board.
- PCI Express riser cards plug directly into the motherboard, allowing PCI Express cards to be installed.
- Two XAUI riser cards provide slots that access to the on-chip 10 Gb Ethernet interfaces on the UltraSPARC T2 Processor. Alternately, these slots can provide access to PCI Express interfaces. Each slot can either accept an optical/copper XAUI card, or an industry standard low-profile PCI Express card with up to an x8 form factor edge connector (x4 electrically). Cards are installed in a horizontal orientation.
- The disk backplane mounts to the disk cages in the two chassis, delivering disk data through one or two 4-channel discrete mini-SAS cables from the motherboard. A 4-disk backplane is offered for the Sun SPARC Enterprise T5120 server while the Sun SPARC Enterprise T5220 server supports an 8-disk backplane.
- A front USB panel card inserts directly into the disk backplane, providing two USB connections to the front of the system.

Sun SPARC Enterprise T5120 Server Overview

The compact Sun SPARC Enterprise T5120 server provides significant computational power in a space-efficient low-power 1U rackmount package. With high levels of price/performance and a low acquisition cost, this server is ideally suited to the delivery of horizontally-scaled transaction and Web services, and can function as a very capable HPC compute node. The server is designed to address the challenges of modern datacenters with greatly reduced power consumption and a small physical footprint. Depending on the model selected, the Sun SPARC Enterprise T5120 server features a single four-, six-, or eight-core UltraSPARC T2 processor.

Enclosure

The 1U Sun SPARC Enterprise T5120 server enclosure is designed for use in a standard 19-inch rack (Table 3).

Table 3. Dimensions and weight of the Sun SPARC Enterprise T5120 server

Dimension	U.S.	International
Height	1.75 inches (1 RU)	44.45 millimeters
Width	18 inches	457.2 millimeters
Depth	28 inches	711.2 millimeters
Weight (approximate, without PCI cards or rackmounts)	20.5 pounds	9.3 kilograms

The Sun SPARC Enterprise T5120 server includes the following major components:

- An UltraSPARC T2 processor with four, six, or eight cores at speeds of 1.2 or 1.4 GHz
- Up to 64 GB of memory in 16 Fully Buffered Dual Inline Memory Module (FBDIMM) slots (1 GB, 2 GB, and 4 GB FBDIMMs supported)
- Four on-board 10/100/1000 Mbps Ethernet ports
- Dedicated low-profile PCI Express slot (x8)
- Two combination XAUI or low-profile PCI Express x4 slots
- Four USB 2.0 ports (2 forward, 2 rear facing)
- Four available disk drives supporting SAS commodity disk drives
- Integrated Lights out Management (ILOM) system controller
- Two (N+1) hot-swappable high-efficiency 650 watt AC/DC power supply units
- Four fan assemblies (each with two fans) populated of a possible eight, under environmental monitoring and control, N+1 redundancy. Fans are accessed through a dedicated top panel door.

Front and Rear Perspectives

Figure 10 illustrates the front and rear panels of the Sun SPARC Enterprise T5120 server.

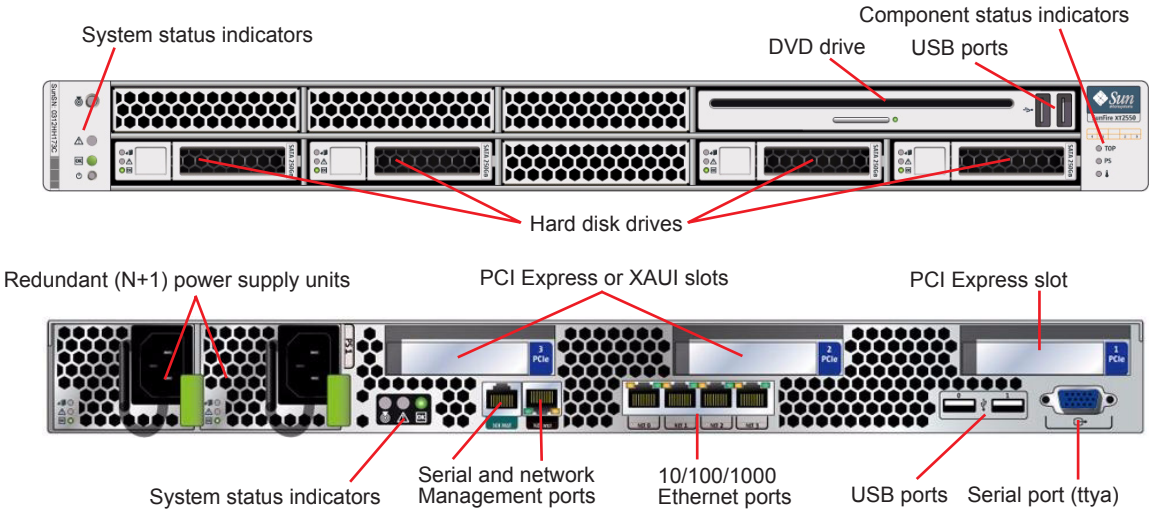

Figure 10. Sun SPARC Enterprise T5120 server, front and rear panels

External features of the Sun SPARC Enterprise T5120 server include:

- Front and rear system and component status indicator lights provide locator (white), service required (amber), and activity status (green) for the system.
- Four hot-plug SAS disk drives insert through the front panel of the system.
- One slimline, slot-accessible DVD-R is accessed through the front panel.
- Four USB 2.0 ports are provided, two on the front panel, and two on the rear.
- Two hot-plug/hot-swap (N+1) power supplies with integral fans insert from the rear.
- Rear power-supply indicator lights convey the status of each power supply.
- A single AC plug is provided on each hot-plug/hot-swap power supply.
- Four 10/100/1000Base-T autosensing Ethernet ports are provided.

- A DB-9 TTYA serial port is provided for serial devices (not connected to the ILOM system controller serial port).
- A total of three PCI Express card slots are provided, two of which can alternately support XAUI cards connected to the UltraSPARC T2 10 Gb Ethernet interfaces.
- Two management ports are provided for use with the ILOM system controller. The RJ-45 serial management port provides the default connection to the ILOM controller. The network management port supports an optional RJ-45/10/100Base-T connection to the ILOM system controller.

Sun SPARC Enterprise T5220 Server Overview

The expandable Sun SPARC Enterprise T5220 server is optimized to deliver transaction and Web services, including Java 2 Platform, Enterprise Edition (J2EE™ platform) technology application services, enterprise application services (ERP, CRM, and SCM), and distributed databases. With considerable expansion capabilities and integrated virtualization technologies, the Sun SPARC Enterprise T5220 server is also an ideal platform for consolidated Tier-1 and Tier-2 workloads.

Enclosure

The Sun SPARC Enterprise T5220 server features a compact, yet expandable 2U rackmount chassis (Table 4), giving organizations the flexibility to scale their processing and I/O needs without wasting precious space.

Table 4. Dimensions and weight of the Sun SPARC Enterprise T5220 server

Server/Dimension	U.S.	International
Height	3.44 inches (2 RU)	87.37 millimeters
Width	16.75 inches	425.45 millimeters
Depth	28.12 inches	714.24 millimeters
Weight (without PCI cards or rack mounts)	40 pounds	28 kilograms

The Sun SPARC Enterprise T5220 server includes the following major components:

- An UltraSPARC T2 processor with four, six, or eight cores
- Up to 64 GB of memory in 16 Fully Buffered Dual Inline Memory Module (FBDIMM) slots (1 GB, 2 GB, and 4 GB FBDIMMs supported)
- Four on-board 10/100/1000 Mbps Ethernet ports
- Four dedicated low-profile PCI Express slots
- Two combination XAUI or low-profile PCI Express x4 slots
- Four USB 2.0 ports (2 forward, 2 rear facing)
- Up to eight available disk drives supporting SAS commodity disk drives
- Integrated Lights out Management (ILOM) system controller
- Two (N+1) hot-plug/hot-swap high-efficiency 750 watt power supplies
- Three fan assemblies (each with two fans) populated of a possible six, under environmental monitoring and control, N+1 redundancy

Front and Rear Perspectives

Figure 11 illustrates the front and back panels of the Sun SPARC Enterprise T5220 server.

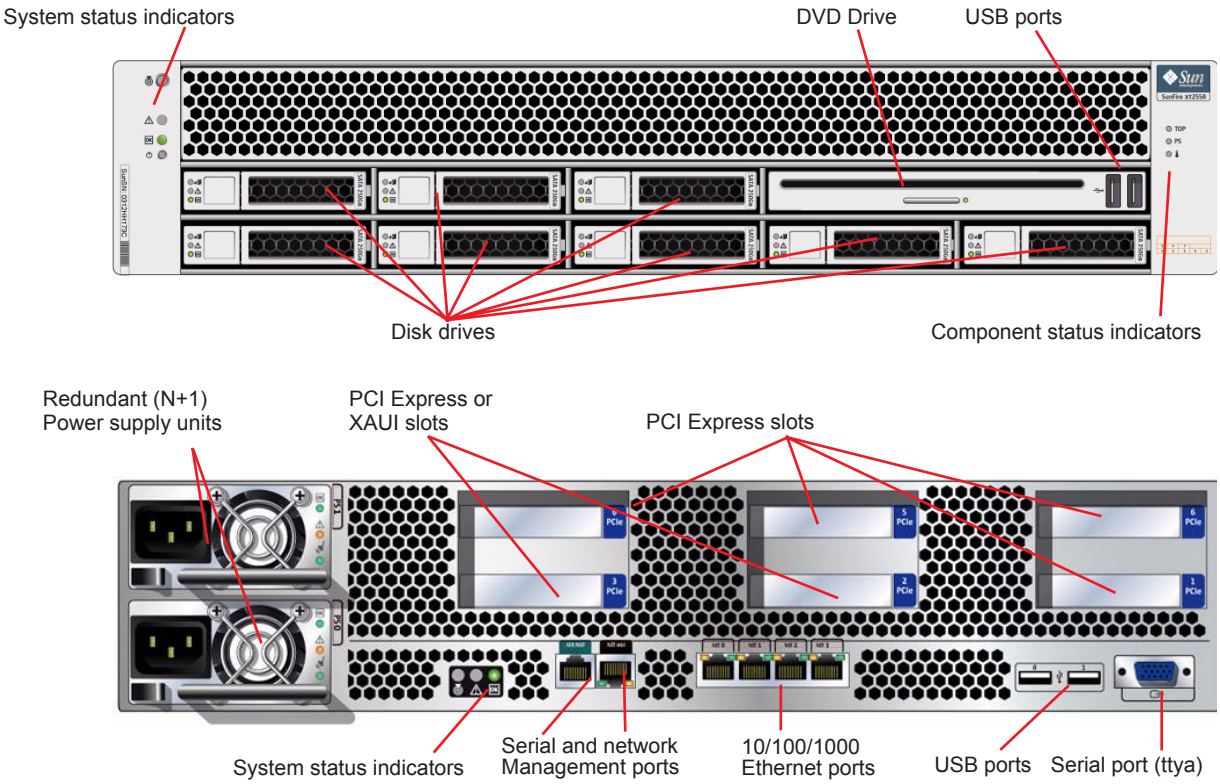

Figure 11. Sun SPARC Enterprise T5220 server, front and rear panels

External features of the Sun SPARC Enterprise T5220 server include:

- Front and rear system and component status indicator lights provide locator (white), service required (amber), and activity status (green) for the system
- Eight hot-plug SAS disk drives insert through the front panel of the system
- One slimline DVD-R drive is accessed through the front panel
- Four USB 2.0 ports are provided, two on the front panel, and two on the rear
- Two hot-plug/hot-swap N+1 power supplies with integral plugs and fans insert from the rear (rear power-supply indicator lights convey the status of each power supply)
- Four 10/100/1000Base-T autosensing Ethernet ports are provided
- A DB-9 TTYA serial port is provided for serial devices (not connect to the ILOM system controller serial port)
- A total of six PCI Express card slots are provided, two of which can support XAUI cards connected to the UltraSPARC T2 10 Gb Ethernet interfaces
- Two management ports are provided for use with the ILOM system controller. The RJ-45 serial management port provides the default connection to the ILOM controller. The network management port supports an optional RJ-45 10/100Base-T connection to the ILOM system controller.

System Management Technology

As the number of systems grow in any organization, the complexities of managing the infrastructure throughout its lifecycle becomes increasingly difficult. Effective system management requires both integrated hardware that can sense and modify the behavior of key system elements, as well as advanced tools that can automate key administrative tasks.

Integrated Lights-Out Management (ILOM) System Controller

Provided across many of Sun's x64 servers, the Integrated Lights Out Management (ILOM) service processor acts as a system controller, facilitating remote management and administration of Sun SPARC Enterprise T5120 and T5220 servers.The service processor is fully featured and is similar in implementation to that used in other Sun modular and rackmount x64 servers. As a result, Sun SPARC Enterprise T5120 and T5220 servers integrate easily with existing management infrastructure.

Critical to effective system management, the ILOM service processor:

- Implements an IPMI 2.0 compliant services processor, providing IPMI management functions to the server's firmware, OS and applications, and to IPMI-based management tools accessing the service processor via the ILOM Ethernet management interface, providing visibility to the environmental sensors (both on the server module, and elsewhere in the chassis)
- Manages inventory and environmental controls for the server, including CPUs, DIMMs, and power supplies, and provides HTTPS/CLI/SNMP access to this data
- Supplies remote textual console interfaces,
- Provides a means to download upgrades to all system firmware

The ILOM service processor also allows the administrator to remotely manage the server, independent of the operating system running on the platform and without interfering with any system activity. ILOM can also send e-mail alerts of hardware failures and warnings, as well as other events related to each server. The ILOM circuitry runs independently from the server, using the server's standby power. As a result, ILOM firmware and software continue to function when the server operating system goes offline, or when the server is powered off. ILOM monitors the following Sun SPARC Enterprise 5120 and 5220 server conditions:

- CPU temperature conditions
- Hard drive presence
- Enclosure thermal conditions
- Fan speed and status
- Power supply status
- Voltage conditions
- Solaris watchdog, boot time-outs, and automatic server restart events

Sun Management Center Software

Sun Management Center software is an element management system for monitoring and managing the Sun environment. Sun Management Center software integrates with the leading enterprise management systems to provide customers with a unified management infrastructure. The base package is free and provides hardware monitoring. Advanced applications (add-ons) extend the monitoring capability of the base package. Sun Management Center software provides:

- Agents for managing Solaris OS (SPARC and x64/x86 platforms) and Linux operating systems
- In-depth hardware and software diagnostics
- Aggregate CPU utilization reporting
- Event and alarm management for thousands of attributes
- Corrective action automation through scripts triggered by alarm thresholds
- Secure management controls for remote dynamic reconfiguration
- The ability to customize modules with a powerful, easy-to-use GUI

Sun N1™ System Manager

The Sun N1™ System Manager is infrastructure lifecycle management software for deploying, monitoring, patching, and managing large and small installations of Sun systems. Sun N1 System Manager takes a step-by-step approach to unraveling the challenges of getting systems operational quickly:

- ***Discover***

  As systems are added to the management network, administrators can use Sun N1 System Manager to discover bare metal systems based on a given subnet address or IP range.

- ***Group***

  Given the number of systems to manage and the constant re-purposing of systems, it is critical for IT organizations to find ways to group resources together. Sun N1 System Manager enables users to logically group systems together and perform actions across a group of systems as easily as performing actions on a single system. Systems can be grouped by function (Web servers versus grid computing), administrative responsibility, or other categorization based on organizational needs.

- ***Provision***

  Sun N1 System Manager remotely installs operating systems (Solaris OS, RedHat, or SuSE Linux) onto selected systems. Administrators can use this functionality to provision operating systems onto bare metal systems or reprovision existing systems. As the infrastructure life cycle continues, Sun N1 System Manager can update firmware and provision software packages and patches to selected systems.

- ***Monitor***
  When systems are up and running, administrators can use Sun N1 System Manager to monitor system health, helping to ensure that everything is running at the optimal levels. The software provides detailed hardware monitoring for attributes such as fans, temperature, disk, and voltage usage, including bare metal systems. Sun N1 System Manager also monitors OS attributes such as swap space, CPU, memory, and file systems. Administrators can define specific threshold levels and set preferred notification methods, including e-mail, pager, or Simple Network Management Protocol (SNMP) traps, for each monitored component as business needs demand.
- ***Manage***
  Businesses require that infrastructure life cycle management extend beyond just deploying and monitoring systems. Sun N1 System Manager includes Lights Out Management capabilities, such as powering systems on and off, and remote serial console access to help IT organizations manage their IT infrastructure from remote locations. Leveraging Sun N1 System Manager software's Role-Based Access Control (RBAC) feature, organizations can grant permissions to specific users to perform specific management tasks.
- ***Hybrid User Interface***
  Sun N1 System Manager offers users a hybrid user interface (UI), accessible from the Web, that integrates both the GUI and CLI into one console. With this hybrid UI, operations performed in the GUI are simultaneously reflected in the CLI, and vice versa.

Chapter 4
Enterprise-Class Software

New technology often requires time for tools and applications to arrive, and delivering agile and highly-available services that take advantage of available resources requires stable development tools, operating systems, middleware and management software. Fortunately, in spite of the breakthrough UltraSPARC T2 processor technology, Sun SPARC Enterprise T5120 and T5220 servers provide full binary compatibility with earlier SPARC systems and are delivered ready to run with pre-loaded tools and the solid foundation of the Solaris OS. Moreover, these systems are provided with a wealth of sophisticated tools that let organizations develop and tune applications as they consolidate and manage workloads while effectively utilizing the resources of the UltraSPARC T2 processor.

Scalability and Support for CoolThreads Technology

The Solaris 10 Operating System is specifically designed to deliver the considerable resources of UltraSPARC T2 processor based systems. In fact, the Solaris 10 OS provides key functionality for virtualization, optimal utilization, high availability, unparalleled security, and extreme performance for both vertically and horizontally scaled environments. The Solaris 10 OS runs on a broad range of SPARC and x86-based systems and compatibility with existing applications is guaranteed.

One of the most attractive features of systems based on the UltraSPARC T2 processor is that they appear as a familiar SMP system to the Solaris OS and the applications it supports. In addition, the Solaris 10 OS has incorporated many features to improve application performance on CMT architectures:

- ***CMT Awareness***
  The Solaris 10 OS is aware of the UltraSPARC T2 processor hierarchy so that the scheduler can effectively balance the load across all the available pipelines. Even though it exposes the UltraSPARC T2 processor as 64 logical processors, the Solaris OS understands the correlation between cores and the threads they support, and provides a fast and efficient thread implementation.

- ***Fine-Granularity Manageability***
  For the UltraSPARC T2 processor, the Solaris 10 OS has the ability to enable or disable individual cores and threads (logical processors). In addition, standard Solaris OS features such as processor sets provide the ability to define a group of logical processors and schedule processes or threads on them.

- ***Binding Interfaces***
  The Solaris OS allows considerable flexibility in that processes and individual threads can be bound to either a processor or a processor set, if required or desired.

- ***Support for Virtualized Networking and I/O, and Accelerated Cryptography***
  The Solaris OS contains technology to support and virtualize components and subsystems on the UltraSPARC T2 processor, including support for the on-chip 10 Gb Ethernet ports and PCI Express interface. As a part of a high-performance network architecture, CMT-aware device drivers are provided so that applications running within virtualization frameworks can effectively share I/O and network devices. Accelerated cryptography is supported through the Solaris Cryptographic framework.

- ***Solaris ZFS File System***
  Solaris ZFS offers a dramatic advance in data management, automating and consolidating complicated storage administration concepts and providing unlimited scalability with the world's first 128-bit file system. ZFS is based on a transactional object model that removes most of the traditional constraints on I/O issue order, resulting in dramatic performance gains. ZFS also provides data integrity, protecting all data with 64-bit checksums that detect and correct silent data corruption.

- ***A Secure and Robust Enterprise-Class Environment***
  Best of all, the Solaris OS doesn't require arbitrary sacrifices. The Solaris Binary Compatibility Guarantee helps ensure that existing SPARC applications continue to run unchanged on UltraSPARC T2 platforms, protecting investments. Certified multi-level security protects Solaris environments from intrusion. Sun's comprehensive Fault Management Architecture means that elements such as Solaris Predictive Self Healing can communicate directly with the hardware to help reduce both planned and unplanned downtime. Effective tools such as DTrace help organizations tune their applications to get the most of the system's resources.

End-to-End Virtualization Technology

Virtualization technology is increasingly popular as organizations strive to consolidate disparate workloads onto fewer more powerful systems, while increasing utilization. Sun SPARC Enterprise T5120 and T5220 servers are specifically designed for virtualization, providing very fine-grained division of multiple resources — from processing to virtualized networking and I/O. Most importantly, Sun's virtualization technology is provided as a part of the system, not an expensive add-on.

A Multithreaded Hypervisor

Like the UltraSPARC T1 processor, the UltraSPARC T2 processor offers a multithreaded hypervisor — a small firmware layer that provides a stable virtual machine architecture that is tightly integrated with the processor. Multithreading is crucial, since the hypervisor interacts directly with the underlying chip-multithreaded UltraSPARC T2 processor. This architecture is able to context switch between multiple threads in a

single core, a task that would require additional software and considerable overhead in competing architecture.

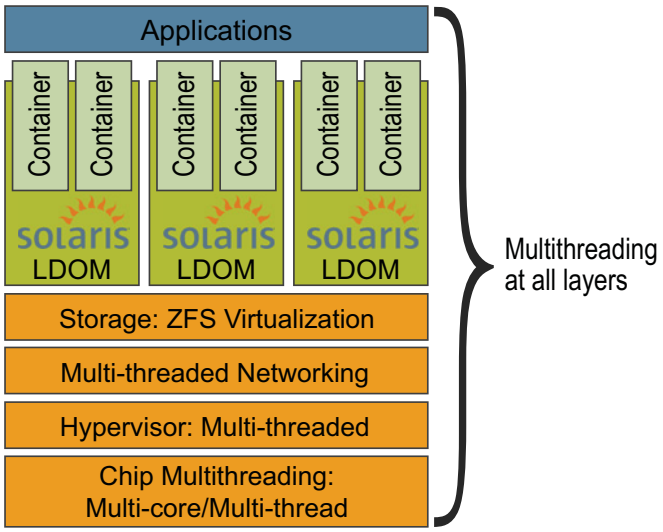

Figure 12. Sun provides parallelization and virtualization at every level of the technology stack

Corresponding layers of virtualization technology are built on top of the hypervisor as shown in Figure 12. The strength of Sun's approach is that all of the layers of the architecture are fully multithreaded, from the processor up through applications that use the fully threaded Java application model. Far from new technology, the Solaris OS has provided multithreading support since 1992. This experience has helped to inform technology decisions at other levels, ultimately resulting in a system that parallelizes and virtualizes at every level. In addition to the processor and hypervisor, Sun provides fully multithreaded networking and the fully multithreaded ZFS file system. Sun Logical Domains (LDOMs), Solaris Containers, and multithreaded applications are able to receive exactly the resources they need.

Sun Logical Domains

Supported in all Sun servers utilizing CMT technology, Sun Logical Domains provide full virtual machines that run an independent operating system instance, and contain virtualized CPU, memory, storage, console, and cryptographic devices. Within the Sun Logical Domains architecture, operating systems such as the Solaris 10 OS are written to the hypervisor, which provides a stable, idealized, and virtualizable representation of the underlying server hardware to the operating system in each Logical Domain. Each Logical Domain is completely isolated, and the maximum number of virtual machines created on a single platform relies upon the capabilities of the hypervisor, rather than the number of physical hardware devices installed in the system. For example, the Sun SPARC Enterprise T5220 server with a single UltraSPARC T2 processor supports up to 64 logical domains[1], and each individual logical domain can run a unique OS instance.

1.Though possible, this practice is not a generally recommended.

By taking advantage of Logical Domains, organizations gain the flexibility to deploy multiple operating systems simultaneously on a single platform. In addition, administrators can leverage virtual device capabilities to transport an entire software stack hosted on a Logical Domain from one physical machine to another. Logical Domains can also host Solaris Containers to capture the isolation, flexibility, and manageability features of both technologies. Deeply integrating Logical Domains with both the UltraSPARC T2 processor and the Solaris 10 OS increases flexibility, isolates workload processing, and improves the potential for maximum server utilization.

The Logical Domains architecture includes underlying server hardware, hypervisor firmware, virtualized devices, and guest, control, and service domains. The hypervisor firmware provides an interface between each hosted operating system and the server hardware. An operating system instance controlled and supported by the hypervisor is called a *guest domain*. Communication to the hypervisor, hardware platform, and other domains for creation and control of guest domains is handled by the *control domain*. Guest domains are granted virtual device access via a *service domain* which controls both the system and hypervisor, and also assigns I/O.

To support virtualized networking, Logical Domains implement a virtual Layer 2 switch, to which guest domains can be connected. East guest domain can be connected to multiple *vswitches* and multiple guest domains can also be connected to the same vswitch. Vswitches can either be associated with a real physical network port, or they may exist without an associated port, in which case the vswitch provides only communications between domains within the same server. This approach also gives guest domains a direct communication channel to the network (Figure 13). Each guest domain believes it owns the entire NIC and the bandwidth it provides, yet in practice only a portion of the total bandwidth is allotted to the domain. As a result, every NIC can be configured as demand dictates, with each domain receiving bandwidth on an as-needed basis.

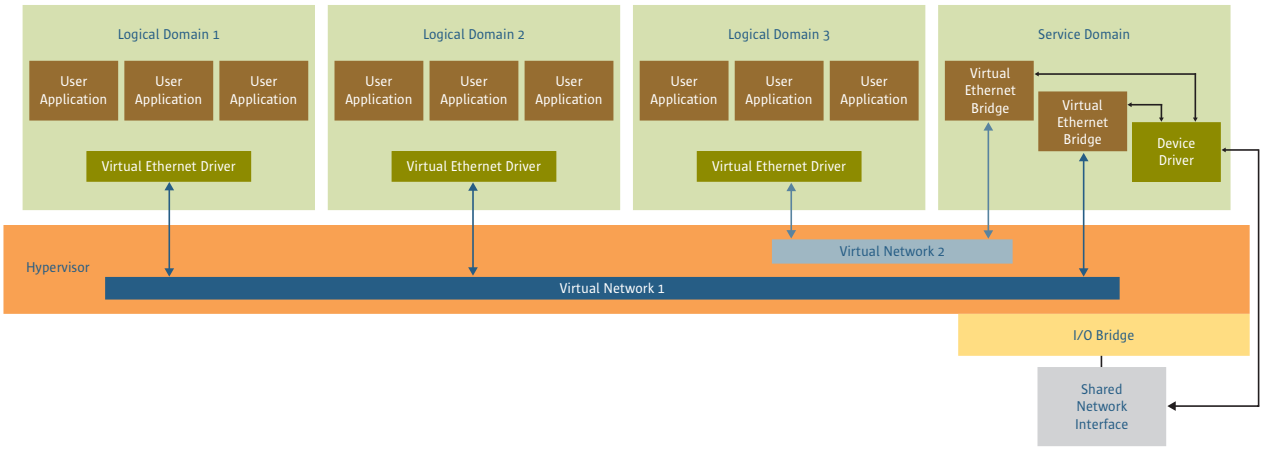

Figure 13. Data moves directly between a Logical Domain and a virtualized device

Solaris™ Containers

Providing virtualization at the OS level, Solaris Containers consist of a group of technologies that work together to efficiently manage system resources, virtualize the environment, and provide a complete, isolated, and secure runtime environment for applications. Solaris containers include important technologies that work together with the fair-share scheduler:

- ***Solaris Zones***

  The Solaris 10 OS provides a unique partitioning technology called Solaris Zones that can be used to create an isolated and secure environment for running applications. A zone is a virtualized operating system environment created within a single instance of the Solaris OS. Zones can be used to isolate applications and processes from the rest of the system. This isolation helps enhance security and reliability since processes in one zone are prevented from interfering with processes running in another zone.

- ***Resource Management***

  Resource management tools provided with the Solaris OS help allocate resources such as CPUs to specific applications. CPUs in a multiprocessor system (or threads in the UltraSPARC T2 processor) can be logically partitioned into processor sets and bound to a resource pool, which in turn can be assigned to a Solaris zone. Resource pools provide the capability to separate workloads so that consumption of CPU resources do not overlap, and also provide a persistent configuration mechanism for processor sets and scheduling class assignment. In addition, the dynamic features of resource pools enable administrators to adjust system resources in response to changing workload demands.

Fault Management and Predictive Self Healing

The Solaris 10 OS introduced a new architecture for building and deploying systems and services capable of fault management and predictive self-healing. Predictive Self Healing is an innovative capability in the Solaris 10 OS that automatically diagnoses, isolates, and recovers from many hardware and application faults. As a result, business-critical applications and essential system services can continue uninterrupted in the event of software failures, major hardware component failures, and even software mis-configuration problems.

- ***Solaris Fault Manager***

  The Solaris Fault Manager facility collects data relating to hardware and software errors. This facility automatically and silently detects and diagnoses the underlying problem, with an extensible set of agents that automatically respond by taking the faulty component offline. Easy-to-understand diagnostic messages link to articles in Sun's knowledge base to clearly guide administrators through corrective tasks that require human intervention. The open design of the Solaris Fault Manager facility also permits administrators and field personnel to observe

the activities of the diagnostic system. With Solaris Fault Manager, the overall time from a fault condition, to automated diagnosis, to any necessary human intervention is greatly reduced, increasing application uptime.

- ***Solaris Service Manager***
  The Solaris Service Manager facility creates a standardized control mechanism for application services by turning them into first-class objects that administrators can observe and manage in a uniform way. These services can then be automatically restarted if they are accidentally terminated by an administrator, if they are aborted as the result of a software programming error, or if they are interrupted by an underlying hardware problem. In addition, the Solaris Service Manager software reduces system boot time by as much as 75 percent by starting services in parallel according to their dependencies. An "undo" feature helps safeguard against human errors by permitting easy change rollback. The Solaris Service Manager is also simple to deploy; developers can convert most existing applications to take full advantage of Solaris Service Manager features by simply adding a simple XML file to each application.

Predictive self healing and fault management provide the following specific capabilities on Sun SPARC Enterprise T5120/T5220 servers:

- *CPU Offlining* takes a core or threads offline that has been deemed faulty. Offlined CPUs are stored in the resource cache and stay offline on reboot unless the processor has been replaced, in which case the CPU is cleared from the resource cache.
- *Memory Page Retirement* retires pages of memory that have been marked as faulty. Pages are stored in the resource cache and stay retired on reboot unless the offending DIMM has been replaced, in which case affected pages are cleared from the resource cache.
- *I/O Retirement* logs errors and faults.
- *fmlog* logs faults detected by the system.

Solaris CoolTools for SPARC: Performance and Rapid Time-to-Market

No matter how compelling new hardware or OS platforms may be, organizations must be assured that the costs and risks of adoption are in line with the rewards. In particular, organizations want to be able to continue to leverage the considerable advantages of popular commercial and open source software. Developers don't want to have to switch compilers and basic development tools. Administrators can scarcely afford a more complex support matrix or more time spent getting applications to run effectively in a new environment. Sun's CoolTools program is designed specifically to take the cost and risk out of moving Web tier environments. CoolTools for SPARC are provided on each Sun SPARC Enterprise T5120 and T5220 server.

Application Selection

Application selection helps identify those applications that stand to benefit from CoolThreads technology. The *CoolThreads Selection Tool (coolst)* helps determine application suitability for both the UltraSPARC T1 and ULtraSPARC T2 architectures, accelerating the understanding of application execution and helping to take the risk out of investment decisions. The tool measures the number of light-weight processes (threads) to determine potential parallelism.

Development

Developers need to be able to build, test, and evaluate applications, producing the most effective code while advancing their productivity with their chosen tools.

- *GCC for SPARC Systems (GCC4SS)* — Specifically tuned and optimized for SPARC systems, GCC4SS complements the popular GCC compiler suite, delivering up to three times the performance of compiled applications with even greater levels of reliability. At the same time, GCC4SS is 100 percent compatible with GCC, supporting all ABIs, language extensions, and flags.
- *Sun Studio 12* — Sun Studio 12 provides developers with the latest record-setting, high-performance, optimizing C, C++, and FORTRAN compiler compilers for the Solaris OS on SPARC and x86/x64 platforms. Command-line tools and a NetBeans-based Integrated Development Environment (IDE) are provided for application performance analysis and debugging of mixed source language applications. In addition to providing multi-platform support, Sun Studio 12 compilers are compatible with GCC, Visual C++, C99, OpenMP, and FORTRAN 2003.
- *Binary Improvement Tool (BIT) and Simple Performance Optimization Tool (SPOT)* — Used for code coverage analysis, BIT provides instruction and call count data at run time, helping to significantly improve developer productivity and application performance. BIT does not require source code, and works with both executables and libraries. SPOT also helps deliver improved developer productivity by automating the gathering and reporting of code data.
- *Sun Memory Error Discovery Tool (Discover)* — Memory access errors can be one of the hardest types of errors to detect, since symptoms of the error typically appear arbitrarily far from the point where the error occurred. The Sun Memory Error Discovery Tool (Discover) is designed to detect and report common memory access errors. Reported errors include accessing uninitialized memory, writing past the end of an array, or accessing memory after it has been freed.

Tuning and Debugging

Administrators and developers alike need to monitor, analyze, and tune applications under real-world conditions. The following tools aid with tuning and debugging:

- *Corestat* — Corestat provides an on-line monitoring tool for core utilization of the UltraSPARC T2 processor, providing a more accurate measure of processor and system utilization than tools that only measure the utilization of individual threads. Implemented as a Perl script and updated for UltraSPARC T2 processor awareness, corestat aggregates instructions executed by all the threads on a single core, revealing the cycles per instruction of key workloads and indicating where more tuning is needed.
- *Automatic Tuning and Trouble-Shooting System (ATS)* — In the interest of automating application tuning, ATS automatically re-optimizes and re-compiles binaries with no need for source code. ATS identifies the inadequate optimization and then automatically rebuilds the application with the correct options for optimization. ATS is a plug-in for GCC4SS and Sun Studio 12.

Deployment

Cool Tools deployment elements provide applications that are already optimized for CoolThreads technology, and save critical time in configuring systems for performance and consolidation. Deployment elements include:

- *CoolTuner* — CoolTuner provides an on-site "virtual" tuning expert, delivering system performance improvements by automatically applying current best practices including both patching and tuning. Depending on administrator experience, CoolTuner can save hours to weeks of effort tuning servers based on Cool Threads technology.
- *Cool Stack* — Cool Stack represents a collection of the most commonly used free and open source applications, pre-optimized for servers based on CoolThreads technology running the Solaris OS. Including such popular applications as Apache, Perl, PHP, Squid, Tomcat, and MySQL, these applications have been recompiled with Sun Studio 12 compilers to deliver a 30 to 200 percent performance improvement over standard binaries compiled with GCC. Cool Stack applications also bring performance benefits to any SPARC system.
- *Consolidation Tool for Sun Fire Servers* — Powerful Solaris Containers offer myriad consolidation possibilities and the Consolidation Tool for Sun Fire Servers speeds their deployment. With a wizard based GUI, this tool simplifies and automates the installation of consolidated applications, enabling even novice administrators to create a fully virtualized and consolidated environment using Solaris Containers. The result is fast and high-quality consolidated deployments.

Sun Java™ Enterprise System (Java ES)

The software industry has traditionally offered point products that solve specific parts of a problem, leaving it to customers to integrate those products into a solution that can support their business applications. Organizations don't purchase their operating systems by assembling core components such as drivers, schedulers, command, and

administration utilities and it doesn't make sense for them to assemble and integrate traditional middleware this way either.

The Sun Java Enterprise System 5 (Java JES 5) provides a complete set of infrastructure software that is integrated to work as a whole, and that offers shared components, common technologies, a consistent architecture and user experience. Using world-class software, Sun redefines the software system from the operating system up through the J2EE specification layer. Customers can write their business applications to Java software standards, leverage Java Enterprise System network services, and Sun delivers the end-to-end solution to run them.

The Java Enterprise System is now available via open source, eliminating costs for both development and deployment. A 90-day Java ES 5 evaluation is provided with each Sun SPARC Enterprise T5120 and T5220 server. The foundation of Java ES 5 is the Java ES 5 Base, including the following products:

- *Access Manager* — Open, standards-based access control, single sign-on and federation services that help control costs and minimize the security risks of conducting business more openly
- *Sun Java System Application Server Enterprise Edition* — A robust, commercial, J2EE 4-compliant application server that makes building robust, scalable enterprise applications easier than ever; and is the perfect platform for implementing SOA and Web 2.0 applications
- *Sun Java System Directory Server Enterprise Edition* — A secure, highly available, scalable, and easy-to-manage directory infrastructure that effectively manages identities in growing and dynamic environments
- *Java Studio Creator* — Rapid visual web application and portlet development
- *Java Studio Enterprise* — Sun Java Studio Enterprise is an award-winning integrated development environment (IDE) for enterprise architects and developers.
- *Message Queue* — Sun Java System Message Queue software is a leading business integration messaging server designed to deliver the exceptional scalability, reliability, and advanced security features necessary for large-scale enterprise deployments
- *Portal Server* — The Sun Java System Portal Server provides a new level of enterprise productivity, enabling users and groups to work together easily and securely within a dynamic organizational structure
- *Solaris Cluster* — Solaris Cluster is a multi-system, multi-site disaster recovery solution that manages the availability of applications services and data across local, regional and vastly dispersed datacenters.
- *Sun Studio* — Sun Studio software delivers high-performance, optimizing C, C++, and Fortran compilers for the Solaris OS on SPARC, and both Solaris and Linux on x86/x64 platforms.

- *Web Server* — As the leading enterprise Web server, Sun Java System Web Server is engineered to meet the stringent requirements of organizations that use Web technologies as a competitive advantage
- *Web Proxy* —The Sun Java System Web Proxy Server is a powerful system for caching and filtering Web content, boosting network performance, and reducing user wait times

Organizations can add to Java ES 5 Base by subscribing to additional services including Sun Java System Identity Manager and the Sun Java Composite Application Platform Suite (Java CAPS).

Chapter 5
Conclusion

Delivering on the demands of Web 2.0 applications and virtualized, eco-efficient data centers requires a comprehensive approach that includes innovative processors, system platforms, and operating systems, along with leading application, middleware, and management technology. With its strong technology positions and R&D investments in all of these areas, Sun is in a unique position to deliver on this vision. Far from futuristic, Sun has effective solutions today that can help organizations cope with the need for performance and capacity while effectively managing space, power and heat.

Building on the successful UltraSPARC T1 processor, the UltraSPARC T2 processor delivers approximately twice the throughput and efficiency, and serves as the industry's first massively-threaded system on a chip. With 64 threads per processor, on-chip memory management, two 10 Gb Ethernet interfaces, PCI express, and on-chip cryptographic acceleration, these systems fundamentally redefine the capabilities of a modern processor. Sun SPARC Enterprise T5120 and T5220 servers leverage these strengths to provide powerful and highly-scalable server platforms while delivering new levels of performance and performance-per-watt in a compact rackmount chassis. The result is datacenter infrastructure that can truly scale to meet new challenges with a very small footprint.

Sun SPARC Enterprise T5120 and T5220 servers provide the computational, networking, and I/O resources needed by the most demanding Web, application, database, and HPC applications, and they facilitate highly-effective consolidation efforts. With end-to-end support for multithreading and virtualization, these systems can consolidate workloads and effectively utilize system resources even as they preserve investments in SPARC/Solaris technology and provide tools for open-source software environments. With innovations such as Sun Logical Domains, Solaris Containers, and Java technology, organizations can adopt these radical new systems for their most important projects — acting responsibly toward the environment and the bottom line.

For More Information

To learn more about Sun products and the benefits of Sun SPARC Enterprise T5120 and T5220 servers, contact a Sun sales representative, or consult the related documents and Web sites listed in Table 5.

Table 5. Related Websites

Web Site URL	Description
sun.com/coolthreads	Sun SPARC Enterprise T5120 and T5220 Servers
sun.com/processors/UltraSPARC-T2	Sun UltraSPARC T2 Processor
opensparc.net/opensparc-t2	OpenSPARC T2
sun.com/processors/throughput	Throughput Computing
sun.com/servers/coolthreads/overview	Sun Servers with CoolThreads Technology
sun.com/servers/coolthreads/ldoms	Sun Logical Domains
sun.com/solaris	The Solaris Operating System
sun.com/software/products/system_manager	Sun N1 System Manager
sun.com/software/products/sunmanagementcenter	Sun Management Center

Sun Microsystems, Inc. 4150 Network Circle, Santa Clara, CA 95054 USA **Phone** 1-650-960-1300 or 1-800-555-9SUN (9786) **Web** sun.com

Printed in USA 10/07 SunWIN 512750